The Fantastic Four: Assessment Models for Sustainable Accountability

Stephanie E. Mikitish

Introduction
What is the value of a college degree? How do university libraries contribute to this value and to the overall education of their students? These important questions are addressed in this paper, which provides an in-depth look at assessment models for academic libraries. To provide some context, one can review research which has demonstrated that earning a college degree benefits both college graduates and society due to graduates’ lower rates of unemployment, better health, and higher earnings, marriage rates, and civic involvement. Consequently, the demand for a college education to earn this demonstrably valuable degree typically remains high. The value of the individual higher education institutions (HEIs) that provide the degrees, however, is not as clear. Faced with calls for accountability from various stakeholders in government and society during an “era of unprecedented change and technological innovation” HEIs must develop ways to conceptualize and measure their value, usually quantitatively. Despite the imperative for accountability, there exist no universally agreed upon indicators of HEI goodness. Even accreditation, viewed by many as the most important federal, regional, or state guarantee of HEI quality, is certainly not perfect. One study found that approximately twenty percent of legal HEIs were unaccredited in thirty-four states surveyed. Additionally, accreditation standards do not include, or only minimally include, many of the units in an HEI including academic libraries.

Academic libraries were one of the earliest support units to develop in HEIs and research indicates that they are still vital to the research and learning activities at the university. However, academic library value, much like an HEI’s value, is not self-evident. Moreover, interviews with HEI administrators found that they believe the roles of academic library roles are changing and that technology has diminished the value of the academic library. A comprehensive study of academic library value by the Association of College and Research Libraries (ACRL) found that most value studies assess an academic library’s use or utility, finances, production of a commodity, impact, and how they compare to competing alternatives. The most significant driver behind changes in academic library research has been the need to acknowledge the audience outside of the library, including its users and parent institutions.

Models are a useful tool in linking otherwise disparate data points into a coherent picture of a library’s worth. However, library and information science (LIS) literature indicates that many academic librarians do not use models in their assessment or evaluative activities, and of the models that do appear in LIS literature, few have been tested or critiqued with the exception of the LibQUAL+ model. Additionally, most publications tend to only evaluate one model or type of model at a time. To date, literature reviews and meta-analyses have been published on gap theory based models, outcomes based models, return on

Stephanie E. Mikitish is a Doctoral Candidate, School of Communication and Information, Rutgers, The State University of New Jersey, e-mail: mikitish@rutgers.edu
investment (ROI) models, \(^{14}\) balanced scorecard models, \(^{15}\) and total quality management (TQM) models. \(^{16}\) This paper describes, compares, and critiques four popular assessment and evaluation models from the LIS literature in the last twenty years, illustrated by studies utilizing them.

**Systems Based Models**

Systems based models were applied in the early days of library assessment and prevalent through the 1980s. Their components include inputs, processes, and outputs (see figure 1). Inputs are the resources that are put into the library, which through various processes become outputs, which are the services or work that the library has done. The push to develop and use these models came from the understanding that the types of information commonly collected for library statistics, such as collection size, were not sufficient indicators of a library’s value. This sentiment also marked the beginning of the shift from standards based value to performance and evaluation based value. Most commonly, studies based on this model focus on a library’s efficiency or monetary value, such as ROI studies.

Variations of the systems model include one by Orr, who developed it in 1973 (see figure 2). Orr’s model is also useful because it separates the concepts of quality, which is related to the library’s capability to provide resources and services, and value, which is related to the library’s beneficial effects on its users and parent institution. \(^{17}\) Despite its usefulness, no studies from the last two decades have utilized this particular variation of the systems model.

**Extensions of Systems Based Models—Inputs and Outputs**

Most of the few studies that extend the systems approach from the last twenty years utilize outcomes based models, which highlight an emphasis on outcomes rather than outputs. Conceptually outputs differ from outcomes in that the former are related to and of importance to library insiders because they are what the library has been able to do. The latter, however, are what library users are able to accomplish through the services and resources that the library provides. It follows from this that one major benefit of systems models are that they tend to use data already collected by several libraries, which lend them to inter-institutional comparisons.
Many of the earlier studies of library value that are grounded in inputs, such as financial information, utilize this model. These include cost-analysis,\textsuperscript{19} contingent valuation,\textsuperscript{20} and value-in-use studies.\textsuperscript{21} For example, an early cost-analysis study by Abels, Kantor, and Saracevic\textsuperscript{22} had a major finding that unit costs for providing resources and services varied greatly between and even within libraries. The two explanations that they gave for the disparities were that the costs for providing resources and services could total many times more from one library to another and the fact that most libraries were working on annual budgets that would be reduced for the next year if the present year’s amount was not spent.\textsuperscript{23} Aabø conducted a meta-analysis of thirty-eight ROI studies, and concluded that the average ROI for the libraries studied was between four and five dollars for every dollar spent by the library.\textsuperscript{24}

A major critique of systems models is that although they can demonstrate correlations between inputs, processes, and outputs they often are not detailed enough to isolate the impact of library factors. For instance, Frade and Washburn’s\textsuperscript{25} study compared the results of a previous and more extensive study from 2001-2002 with comparable usage data from 2003-2004. Their stated goal was to investigate the possible effect of longer hours and the creation of an information commons, which are both input measures, on library usage, which was their main output measure. While the students in the first study were interviewed and surveyed about why they were using the library, which identified their needs and gathered some qualitative information on whether their needs were being met, the second study only utilized usage statistics, such as gate counts and circulation figures. Therefore, there was no clear link between the change in the inputs or the usage outputs, even though library use increased.

Some researchers have suspected that the relationships between library inputs, processes, and outputs may be more complex than can be explained with a simple linear model. Allen and Dickie\textsuperscript{26} researched the relationship between library funding and the demand for library services. Using Association of Research Libraries (ARL) data for 113 members over eleven years they then ran a multiple regression with demand for services and demographic institutional factors as independent variables and library funding as the dependent variable. Their results suggested a logarithmic model for funding, rather than a linear one, because it would diminish the impact of a change in demand for services.\textsuperscript{27} Another research team used self-organizing map cluster analysis to examine the relationships among measures of resources and service usage.\textsuperscript{28} This study used ACRL metric data from 1,395 US and Canadian libraries and created a library performance metric, which was a composite of circulation, instruction session, and reference data. A significant finding from this study was that libraries with low monetary, personnel, and material resources could still have high performance metrics if they offered higher service features such as giving more presentations to groups, being open more hours, and staffing more service points.\textsuperscript{29}

Although systems based models are useful because they focus on library performance, their analysis ends at outputs that are useful for the library and do not fully demonstrate the library’s value to outside entities. Thus it is unclear whether or how library users benefit from the library’s resources and services. This is why there are few studies in the last twenty years that exclude outside perspectives, namely those of library users and the academic library’s parent institution. Outcomes based models developed to include these perspectives.

**Outcomes Based Models**

Student outcomes are defined as results or competencies that students gain from their college education. They did not gain prominence until the early 1990s, but in the last twenty years most studies of value in HEIs or their subordinate units attempt to demonstrate their impact on this type of metric. By focusing on student or other user related outcomes libraries are able to demonstrate their value from the perspective of individual users and their parent institutions. Outcomes based models extend earlier system based...
models with outcomes following outputs (see figure 3). Variations of the model include ones by Fraser, McClure, and Leahy\textsuperscript{41} and Rubin,\textsuperscript{32} which have not been utilized in any published studies, and one by Markless and Streifield that has been utilized in published studies.\textsuperscript{33} The latter includes the concept of aggregated outcomes, which are comprised of individual student outcomes that are of interest to the parent institution.\textsuperscript{44} Recent ROI studies also utilize this model.\textsuperscript{35}

**Studies Utilizing Outcomes Based Models**

Outcomes based studies are most easily broken into groups by the types of outcomes that higher education researchers are interested in, such as retention,\textsuperscript{36} persistence,\textsuperscript{37} graduation,\textsuperscript{38} and achievement.\textsuperscript{39} Learning\textsuperscript{40} related outcomes for students include critical thinking,\textsuperscript{41} engagement,\textsuperscript{42} and satisfaction with their educational experience.\textsuperscript{43} Outcomes related to university faculty or researchers include teaching,\textsuperscript{44} productivity,\textsuperscript{45} institutional ranking,\textsuperscript{46} and academic reading.\textsuperscript{47}

Gann and Pratt’s study focused on service usage, but because they took into account user outcomes their study was based on the outcomes model rather than the systems one.\textsuperscript{48} Their goal was to demonstrate how the number of hours of research assistance they provided for a medical research library at the University of Texas contributed to requester outcomes, such as publication and non-patient education. Although they were able to demonstrate how these factors impacted user outcomes,\textsuperscript{49} there was no link between the resources provided by the library and their output, or how their outcomes looped back and influenced the library’s inputs. This type of assessment highlights one of the weaknesses of solely comparing library inputs to outputs or even outcomes. It does not take into account any of the library’s actions that could increase or decrease impact, and if the library’s users simply use the service less the library’s impact will decrease. Moreover it does not take quality into account. If a librarian spends more time than usual on a request and does an exemplary job but the user does not get published, then it will decrease the library’s impact and may make the individual librarian seem less efficient.

The most significant issue with using outcomes assessment is that this type of assessment uses input data from libraries and outcome data related to students or other users, but it often does not take into account student related input data, such as demographics. Astin, one of the education field’s earliest and most significant supporters of studying student outcomes has stated in the *Chronicle of Higher Education* that one can predict a student’s ability to graduate sixty-seven percent of the time based on incoming demographic variables alone,\textsuperscript{50} which means that college outcomes largely depend on the students who enroll, not the institution.\textsuperscript{51} This prediction can lead to potential issues with the validity of a study’s findings. For instance, a study by Wong and Cmor studied the relationship between student GPA and the number of library information literacy sessions they attended.\textsuperscript{52} Although they found that higher GPAs were increasingly positively associated with the number of library workshops attended, there was only one strong association where one student had attended five workshops.\textsuperscript{53} Because the confounding effect of other variables was not separated from the library variables this study does not necessarily show the library’s impact on GPA. It is possible, for instance, that the students may have had higher GPAs even without the library sessions due to their having higher intelligences or being enrolled in different majors than those attending no or fewer sessions.

Another related issue with outcomes assessments is that academic libraries tend to have an indirect ef-
fect on student education, and consequently a diminished effect on learning or other outcomes.\textsuperscript{54} Kuh and Gonyea studied how the academic library promoted student engagement in learning.\textsuperscript{55} Based on data from over 300,000 students from 1984-2002 from the College Student Experiences Questionnaire (CSEQ) they were able to control for the demographic input variables of students and institutions. They also developed three outcome measures for gains in information literacy, overall gains in college, and satisfaction with the college experience. Although their study did not find that any library experiences directly contributed to any of the outcomes, they explained that this was probably due to the questions on the CSEQ not being meant to measure information literacy.\textsuperscript{56}

Whitmire utilized Astin's Inputs-Environment-Output (I-E-O) model (see figure 4 in the Appendix) in what was the only LIS study found to utilize this variation of the outcomes model.\textsuperscript{57} Despite the model's name, the outputs studied were student outcomes, and she found that student background characteristics (e.g., grades and class year); library experience (e.g., focused library activities); experiences with faculty (e.g., informal interactions); course learning experiences (e.g., active course learning); and writing experiences (e.g., conscientious writing) were the most influential on a student's critical thinking.\textsuperscript{58} This study demonstrates that it is possible for outcomes based models to account for the effect of non-library variables on student outcomes, but many studies do not. A similar issue exists for studies that utilize service quality based models.

**Service Quality Based Models**

The service quality based models used by libraries are informed by gap theory, which is based on identifying discrepancies between expectations and perceptions of service.\textsuperscript{59} The first instrument developed using gap theory was the SERVQUAL instrument designed by Zeithaml, Parasuraman, and Berry. The SERVQUAL survey asks questions based on five dimensions, which include tangibles, reliability, responsiveness, assurance, and empathy. For each item there can be no gap, or a positive or negative gap between an individual's expected minimum level of service and their perceived level of service. Because minimum level of service is subtracted from the perceived level of service, a negative gap score indicates that the individual was not satisfied with the service and a positive gap score indicates that the individual was satisfied with the service.

The ARL commissioned the creation of the LibQUAL+ survey instrument in the 1990s as part of their New Measures Initiative, and it was first issued in 1999. Cook, Heath, and Thompson, from the University of Texas, developed the survey, with input from ARL, particularly Kyrillidou, who would later develop and test the LibQUAL+ Lite survey instrument. Through 2012, LibQUAL+ has been used in over 1,200 various types of libraries worldwide to collect over a million responses.\textsuperscript{60} This makes it one of the most used library assessment tools,\textsuperscript{61} and in academic library literature reviews of assessments it is often the only tool given its own section.\textsuperscript{62} The studies using LibQUAL+ can focus on the instrument itself, but they most often focus on its application in academic libraries.

**Studies Utilizing Service Based Quality Models**

In the academic library literature, Nitecki was an early proponent of SERVQUAL,\textsuperscript{63} but later studies concluded that SERVQUAL was not directly applicable to academic libraries.\textsuperscript{64} This is what drove ARL

---

**FIGURE 4**

Astin’s I-E-O Model. Adapted from: Joseph R. Matthews, The Evaluation and Measurement.\textsuperscript{65}
to commission the research that led to the development of LibQUAL+. The earliest works on this instrument deal with its development and testing, which is covered in great detail in an earlier critique of the model.\(^6\) Development began when Cook and Heath conducted sixty interviews of students and professors at various levels to identify themes of quality library services, which were found to be ubiquity and ease of access, library as place, and self-reliance.\(^6\) These were the original dimensions that created the service based model, which now consists of Affect of Service, Information Control, and Library as Place dimensions\(^6\) (see figure 5).

Testing of the instrument has found the three-factor model to be valid\(^6\) and psychometrically stable,\(^7\) but a few studies have questioned and tested various aspects of the model. Roszkowski, Baky, and Jones investigated whether perceived score or superiority gap score, which is the score between the perceived and desired levels of service, was better at measuring user satisfaction.\(^7\) They found that perceived score was a more valid indicator of satisfaction, and that the superiority gap score nearly disappeared when the perceived score was partialed out, which caused them to question the use of LibQUAL+ as a measurement of service quality.\(^7\) In other words, service quality is measured by the superiority gap score, but their respondents were judging it by their perceived levels of service. One potential weakness of this study is that it only focused on one institution, but this may have been due to a limited access to data rather than a flaw in study design.

Two additional studies critique the models utilized by SERVQUAL and LibQUAL+. Shi and Levy evaluated SERVQUAL, disconfirmation theory, and LibQUAL+.\(^7\) Their major critiques of the latter were that the response rates are typically lower than the recommended ten percent, and that adding the minimum level of service to SERVQUAL’s expected (“desired” in LibQUAL+ terminology) and perceived levels problematized the concept of service quality. This is because a positive gap between perceived and minimum levels of service was considered an indicator of good quality, even if the perceived level of ser-

---

**FIGURE 5**

LibQUAL+ Model. Adapted from: Association of Research Libraries. “LibQUAL+® Tools.”

<table>
<thead>
<tr>
<th>Scores</th>
<th>Gaps</th>
</tr>
</thead>
<tbody>
<tr>
<td>Desired</td>
<td>Superiority</td>
</tr>
<tr>
<td>Perceived</td>
<td>Adequacy</td>
</tr>
<tr>
<td>Minimum</td>
<td></td>
</tr>
</tbody>
</table>

---

Customer Information Resources

Financial Organizational Readiness

Internal Processes

Desired Scores

Perceived Scores

Minimum Scores

Superiority Gaps

Adequacy Gaps

Service Quality

---

AQL 2015
vice was less than the desired level of service. Yu, Hong, Gu, and Wang have published a detailed critique of gap theory, applying it to both SERVQUAL and LibQUAL+. They gathered 756 SERVQUAL based surveys similar to those used by Cook and Thompson and interviewed fifty library users from three different libraries. In addition to learning that some respondents had difficulty quantifying their experiences with library service, they also found the results to vary widely, which they claimed was due to the user’s library and life experiences. This finding led them to recommend that instruments based on gap theory needed to add context to the participants’ responses, and to suggest that in the library context users should be differentiated by library sophistication levels and library differentiated levels of service. Despite these criticisms, LibQUAL+ has been used at hundreds of libraries around the world to gauge levels of library service quality, and to help with strategic planning and benchmarking.

Most studies that utilize the LibQUAL+ instrument and model tend to focus on a specific service, usually at one location. A few researchers studied the same institution over a period of time, or looked at multiple institutions. Initially most researchers used LibQUAL+ to facilitate overall strategic planning, but later studies used it for benchmarking, which suggests a shift in assessment expectations. In 2004 ACRL released its first unified Standards for Libraries in Higher Education. This document focused on what libraries could demonstrate as contributing to student learning outcomes and institutional effectiveness, and it also noted the usefulness of peer comparison, a practice related to benchmarking.

Two studies that utilized this model at one institution were conducted by Knapp and Greenwood, Watson, and Dennis. Knapp summarized how the University of Pittsburgh had used the LibQUAL+ survey over three years to determine satisfaction, and stated it could gauge whether initiatives enacted to address areas of concern were successful based on the next year’s responses. However, a longer longitudinal study by Greenwood, Watson, and Dennis found that after ten years they could not find a direct relationship between library policy changes and mean adequacy gap scores or comments, which they attributed to the data’s complexity, and the fact that the survey was offered every other year, unlike in Knapp’s shorter study.

The last type of applied LibQUAL+ studies covers those between multiple institutions, including consortia. A majority of the studies use LibQUAL+ results for benchmarking, but Hunter and Perret’s study is notable because it also incorporated ACRL statistics, Carnegie Classifications, and funding at seventy-three institutions in a 2007 study on whether higher funded libraries also had the most satisfied users. They found that larger and better funded libraries had users with higher expectations, which meant that their LibQUAL+ measured levels of satisfaction were not necessarily higher than those of smaller, less well funded libraries. Once again this finding underscores the need for context sensitive analyses. Later studies would use other data in order to contextualize the LibQUAL+ results, but most do not find any strong correlations between the data sets.

The obvious strength of service quality based models for academic libraries is that they can incorporate LibQUAL+ to measure user satisfaction. LibQUAL+ is the most widely used assessment instrument of library value due to its psychometric validity and reliability, relative ease of distribution and analysis, and backing by ARL. However, by itself it cannot demonstrate a library’s impact on the institution. Just because users are satisfied with a service does not mean that it is particularly helpful, and therefore valuable to them. At best, LibQUAL+ can indicate potential areas of concern for libraries, and it can measure satisfaction as an outcome of a library service.

Balanced Scorecard Based Models
The balanced scorecard is the only model that lends itself to evaluation rather than assessment because requires stated goals and then tracks whether the goals are met over time. Developed by Kaplan and Norton, it balances evaluations in the areas of users, finances,
internal processes, and potentials. Matthews’ library balanced scorecard is the one most often cited in LIS literature, and it demonstrates how the financial perspective affects the organizational readiness perspective, which affects the internal process perspective, which affects the information resources perspective, which affects the customer perspective (see figure 6). This model utilizes data gathered from the library and users, and it views the results from user, library, and institution perspectives.

Studies Utilizing Balanced Scorecard Based Models
While there are a couple of studies that describe its use at one institution most studies focus on the use of the balanced scorecard at multiple institutions. An example of the former is a study by Krarup, who described how the balanced scorecard helped incorporate internal, user, developmental, and economic perspectives at the Royal Library in Copenhagen over a three year period. One of the earliest studies using the balanced scorecard was performed in 2001 by Poll, who applied it to state and academic libraries in Germany and illustrated how the balanced scorecard guides the management process by allowing the institution to set clear, measurable goals.

One issue with this model is that although academic libraries do similar types of work for the same types of purposes it does not mean that libraries will have comparable measures. This was highlighted by a study done by Mengel and Tolson who compared measures from academic libraries at three different institutions and found that of the total ninety-four measures there was slightly less than a ten percent overlap. ARL is working on creating a common set of benchmarks and measures and there has been a one year pilot project where four different institutions utilized the balance scorecard. More recently Town and Kyrillidou have announced the start of a values based scorecard that can be used in conjunction with a balanced scorecard with dimensions of relational capital, library capital, library virtue, and library momentum.

The most significant advantage to using the balanced scorecard is that it is evaluative and can utilize diverse sets of data. Taylor and Heath describe their assessment efforts at the University of Texas as utilizing data from a diverse range of data, which includes LibQUAL+ and ClimateQUAL results, operating statements, online feedback, and comment cards, and incorporating it into a balanced scorecard that consists of fourteen strategic objectives. The data and the balanced scorecard provide context with which to demonstrate how well the library is meeting its objectives year to year.

Conclusion
The above analysis reviewed the four most frequently used models of assessment of library value. Although each model has its strengths and weaknesses, the LibQUAL+ model has been the most frequently applied framework and the balanced scorecard model seems to best demonstrate qualitative and quantita-
tive impact. Models can illustrate the thought process behind assessments and evaluations. As such they can help to group studies by model, which allows one to compare the results of their study with similar ones. Future models should incorporate measurable goals, such as student outcomes, as illustrated by balanced scorecard based models. They should also incorporate diverse data sets, which could include those collected by the LibQUAL+ survey instrument. These datasets should include non-library data as appropriate, such as student related input data if a study is measuring its impact on student related outcomes. While this paper is limited in that not all studies are published and only those written in the English language were used as examples, its descriptions and critiques of the four most commonly used models of library value can provide a springboard for future studies in academic library value.

Notes
17. Orr, "Measuring the Goodness"
22. Abels et al., "Studying the Cost.”
23. Ibid.: 220.
25. Frade & Washburn, "The University Library”
27. Ibid.
29. Ibid.: 127.
31. Fraser, Bruce T., Charles R. McClure, and Emily H. Leahy. "Toward a Framework for Assessing Library and Institu-


34. Ibid.


42. Kuh & Gonyea, "The Role"

43. Kuh & Gonyea, "The Role"

44. Lindauer, "Defining and Measuring"


49. Ibid.


52. Wong & Cmor, "Measuring Association"

53. Ibid.


55. Kuh & Gonyea, "The Role"

56. Ibid.


58. Ibid.


66. Yu et al., "An Epistemological Critique"

67. Cook, Colleen, and Fred M. Heath. "Users’ Perceptions of
The Fantastic Four: Assessment Models for Sustainable Accountability

71. Roszkowski, Michael J., John S. Baky, and David B. Jones. ”So Which Score on the LibQUAL+™ Tells Me if Library Users are Satisfied?” Library & Information Science Research 27, no. 4 (2005): 424-439.
72. Ibid.
73. Shi & Levy, ”A Theory-guided Approach”
76. Cook & Thompson, “Users Perceptions”
77. Yu et al., ”An Epistemological Critique”
78. Ibid.
85. Knapp, ”We Asked Them”
86. Greenwood et al., ”Ten Years of LibQUAL”
87. Knapp, ”We Asked Them”
88. Greenwood et al., ”Ten Years of LibQUAL”
89. Jankowska et al., ”Expanding Library Service Opportunity”; Lessin, Barton. ”Mining LibQUAL+™ Data for Pointers to Service Quality at Wayne State University.” Journal Of Library Administration 40, no. 3-4 (2004): 139-155.; Hunter & Perret, ”Can Money Buy Happiness”
90. Gatten, ”Measuring Consortium Impact”; Sanville, ”De-
91. Hunter & Perret, “Can Money Buy Happiness”
100. Poll, “Performance, Process, and Costs”
101. Mengel & Lewis, “Collaborative Assessment”
103. Lewis et al., “Building Scorecards”
105. Taylor & Heath, “Assessment and Continuous Planning”